

# #recordsmanagement 2.0

*managing social media as  
corporate record*

[dpconline.org](http://dpconline.org)



Digital **Preservation** Coalition

sara day thomson

tw: @sdaythomson

an introductory workshop

# what is a social media record?

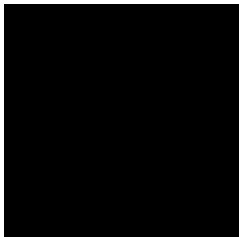
any communication through a  
social media platform:



individual **tweets** or  
conversations (using  
replies) on twitter



**video** or series of  
videos posted to  
youtube



**posts** made on a  
facebook page



**image** uploaded  
to instagram

# why preserve social media records?

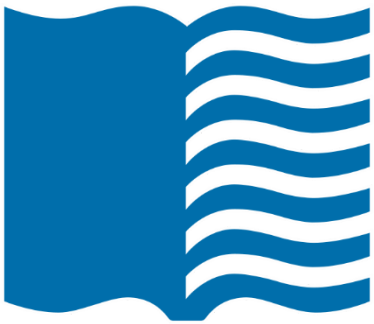
what do you think...?

- \_\_\_\_\_
- \_\_\_\_\_
- \_\_\_\_\_
- \_\_\_\_\_
- . . . .

- compliance and regulation
  - Public Records Act
  - FCA FG15/4
- re-use and new revenues
- marketing and advertising
- corporate history and archives

# selection: defining social media collecting policy

- the ‘conversation’
  - across users
  - across social media services
- embedded media
  - images
  - audio
  - moving images
  - URLs
    - \*shortened links (eg bitly, tinyurl)
- deviations in use
  - lack of uniformity of language, tags, and keywords
  - spelling errors
  - obscenity or libel
  - relevance



# case study: Twitter Research Access at the Library of Congress



## 2010 agreement

- entire archive of tweets from 2006 – 2010
- all on-going tweets
- authenticated research access only
- ... not yet available

## Gov't Accountability Office audit

- no leadership or oversight for information technology
- failure to observe institutional procedures
- bypassing selection process
- lack of a schedule, cost analysis, and scoping

The GAO report (March 2015):

<http://www.gao.gov/assets/670/669367.pdf>

# policy and regulations

- Public Records Act
- Open Data initiatives
- Platform Terms and Conditions
- Intellectual Property Rights
- Data Protection and Ethics

# platform terms and conditions

## common policies

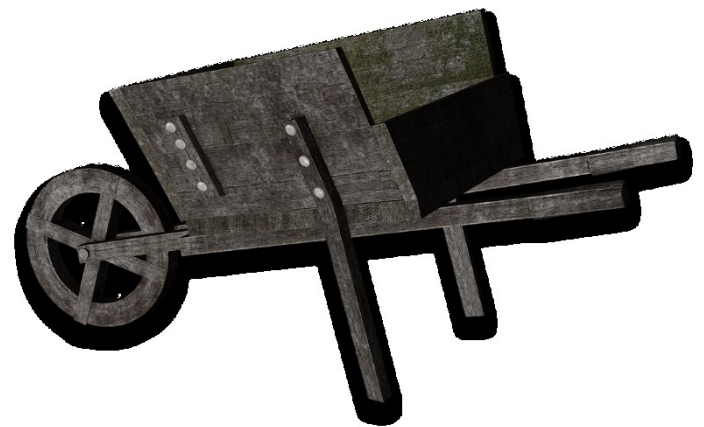
- terms of service
- user agreement
- privacy policy
- terms of use
- developer agreement or developer policy

## developer agreement or policy

- controls use of api
- forbids sharing of data (incl. cloud storage)
- forbids preservation of deleted data

# harvesting: capturing and managing social media records

- platform self-archiving services
- platform application programming interfaces (api)
- os api-based tools
- web crawlers (eg Heritrix)
- data resellers
- third-party services

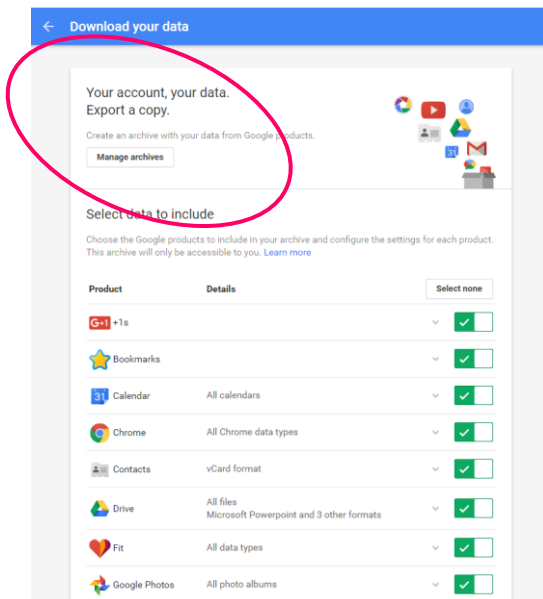


# a quick win demo: platform 'self-archiving'

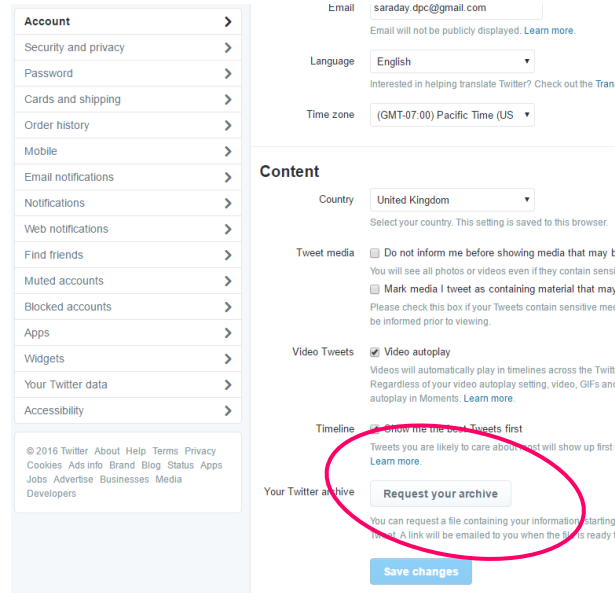


## download your social media

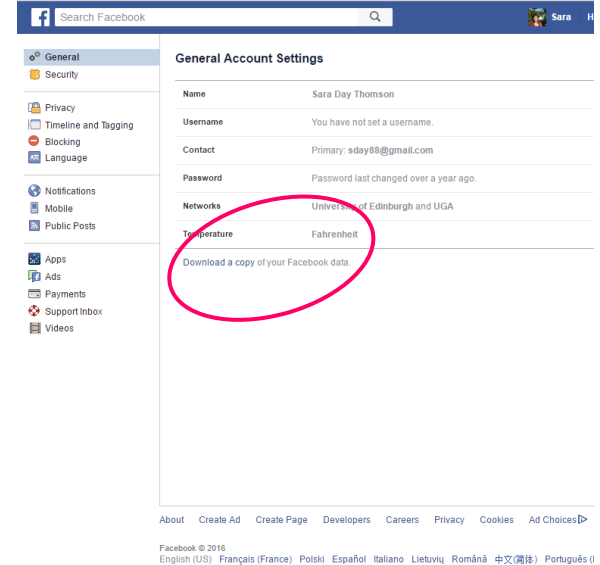
- service for account owner
- structured and unstructured data
- JSON files with metadata



Google



Twitter

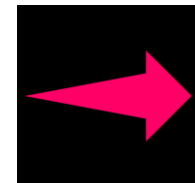


Facebook

# download your tweets

- account settings
- only permitted for the account owner
- good practice for institutions with one or more public-facing social media accounts
- good practice for personal digital archiving

[Go to Twitter](#)



It's Tweet archive time Inbox x



**Twitter** <notify@twitter.com>  
to me ▾

11:14 AM (1 minute ago) ☆



## Sara Day Thomson, your Tweet archive is ready

[Download now](#)

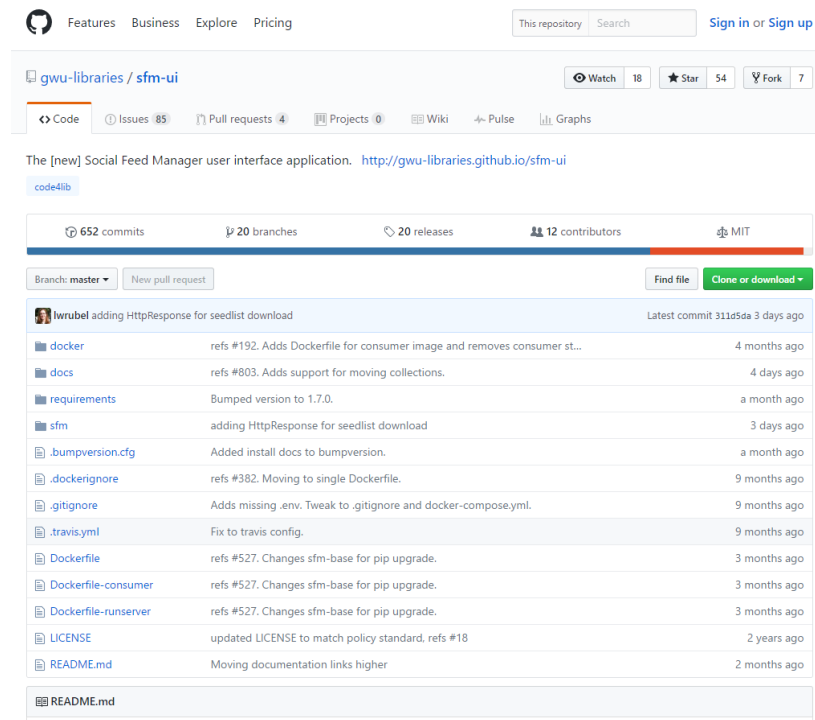
[Need help?](#)

This email was meant for @sdaythomson  
Twitter, Inc. 1355 Market Street, Suite 900 San Francisco, CA 94103

# os api-based tools

## Social Feed Manager (sfm)

- open source tool created by George Washington University Libraries
- captures data from Twitter, Flickr, Sina Weibo, and Tumblr
- captures linked URLs and media
- functionality for managing and searching archived collections
- currently under development, good time to test and give feedback for your own requirements



sign up: <http://ec2-54-87-158-221.compute-1.amazonaws.com/ui/>

read more: <http://gwu-libraries.github.io/sfm-ui/>

# long-term preservation

- vulnerable to loss
- no legal or regulatory requirement for platform to preserve
- changes in platform policy and ownership
- historical data less commercially valuable
- expensive to curate and store



# #thanks!

